Solvation Parameters for Amino Acids

BRIAN J. SMITH

Biomolecular Research Institute, 343 Royal Parade, Parkville, 3052 Victoria, Australia

Received 2 April 1998; accepted 31 October 1998

ABSTRACT: Atomic radii have been derived for the common amino acid side chains using a solvent interaction potential (SIP) based on quantum mechanically derived charges. Solvation energies calculated using these parameters are compared with those obtained using other sets of radii and charges, and from alternative methods. The differences from the experimental solvation energies for the nonionizable residues are all less than 10 kJ mol^{-1}. The largest error in the solvation energy occurs for acetic acid (-16.0 kJ mol^{-1}). For the charged side chain systems the difference from experiment are all less than 10 kJ mol^{-1}. SIP parameters for the aminoacetaldehyde derivatives of the common amino acids are presented. These are used in the calculation of the relative binding energies of six benzamidine inhibitors with trypsin. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 428–442, 1999

Keywords: atomic radii; solvation; amino acids; solvent interaction potential; electrostatics; ligand interaction

Introduction

Electrostatic interactions in macromolecules have long been recognized as important factors influencing their structural and functional properties,[1] and therefore the chemical and physical behavior of these systems is strongly influenced by the dielectric properties of the solvent in which they are immersed. Continuum methods offer one approach to model this effects.[2] These methods generally represent the solvent as a homogeneous dielectric, with the solute located inside a cavity within the solvent dielectric. The shape and size of the cavity are critical to the outcome of these calculations.[3] One of the more successful approaches for computing the electrostatic component to the solvation free energy is the finite difference Poisson–Boltzmann (FDPB) method,[4] which allows the solute cavity to adopt an arbitrary shape. The most widely used method for describing the cavity uses overlapping atom-centered spheres. Each individual atom in the molecule thus requires a radius and partial charge to be assigned. These parameters can be taken directly from molecular mechanics force fields or derived specifically for a particular application.

Correspondence to: B. J. Smith; e-mail: brian.smith@bioresi.com.au

CCC 0192-8651 / 99 / 040428-15

Another very successful alternative for the finite difference method is the boundary element method.[5]

The PARSE parameter set[6] was developed for a range of small polar molecules that are representative of the amino acid side chains and the polypeptide backbone. These were obtained by optimizing 68 charge and 8 radii parameters to a total of 23 chemical groups. These parameters were shown to be extremely successful in predicting solvation free energies; from a selection of 65 molecules, the average error was just 1.8 kJ mol^{-1}, with no error exceeding 8.0 kJ mol^{-1}. For the amino acid side chains the agreement (with the experimental solvation energies known at that time) was even more noteworthy, with no error exceeding 1.0 kJ mol^{-1}. The most severe drawback of this method is that parameters for many atom types are not defined. Thus, for example, attempts to calculate the electrostatic interaction energy of a ligand with a protein could be thwarted if the parameters required to define the ligand are unavailable.

A method has recently been described[7] for determining atomic radii of solute molecules from a solvent interaction potential (SIP). This method relies on locating, on a grid surrounding the solute, the positions of most favorable interaction between the solute and a probe whose characteristics define the solvent of choice.

The potential incorporates both electrostatic and nonelectrostatic interactions. The nonelectrostatic energy is calculated using a Lennard–Jones potential (E_{LJ}), whereas the electrostatic energy is determined from the interaction of the solute partial atomic charges with the dipole moment (E_{i-d}), quadrupole moment (E_{i-q}), and ion-induced dipole (E_{i-id}) of the solvent probe. The total solvent interaction potential, E_{SIP}, is then given by the sum of these components:

$$E_{SIP} = E_{LJ} + E_{i-d} + E_{i-q} + E_{i-id} \quad (1)$$

The van der Waals radii (r_{LJ}) used to describe the LJ potential and a parameter that describes the offset between the center of the probe and the center-of-mass of the probe (Γ) are determined by minimizing the error in the calculated solvation free energy of a set of small molecules, whose solvation free energies are relatively well known. This parameterization set contains both neutral and charged solutes. The gas-phase dipole and quadrupole moments and the offset are scaled by a Langevin function to provide the average effective moments and offset in the direction of the field generated by the solute partial atomic charges. Grid positions about the solute that project a minimum to the nearest atom form the solvent-accessible surface. This surface is used to generate the atomic radii, r_{SIP}.

For the nonelectrostatic component of the free energy of solvation, the following expression was applied:

$$\Delta G_{nonelec} = \gamma A_{SIP} + b \quad (2)$$

where A is the solute solvent-accessible surface area, and γ and b are 0.02 kJ mol^{-1} $Å^{-2}$ and 3.5 kJ mol^{-1}, respectively.[6] The values of γ and b were obtained by fitting this expression to experimental alkane transfer free energies, and are typical for a variety of radius sets used to describe the solute solvent-accessible surface area.[8] The solvent-accessible surface area is obtained from the minima in the solvent interaction potential. The total solvation free energy is then given by the sum of electrostatic and nonelectrostatic components:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{nonelec} \quad (3)$$

In this work, SIP parameters for amino acids have been calculated. The charges and radii for the aminoacetaldehyde analogs for each of the common amino acids are reported. From these, parameter assignment of polypeptides and proteins can be made. Also, the solvation energies for the amino acid side chain molecules are calculated and compared with experiment and with the energies calculated using the PARSE parameters,[6] and those predicted using a recent[9] semiempirical molecular orbital approach (LSQM). The utility of this approach for determining atomic radii and charges is demonstrated through calculations on the relative binding energies of six benzamidine inhibitors of trypsin.

Computational Details

Standard *ab initio* molecular orbital calculations[10] were performed using the GAUSSIAN-94[11] and GAMESS-US[12] programs. Geometries were optimized at the MP2/6-31G(d) level. The geometry about both the amine nitrogen and the aldehyde carbon of the aminoacetaldehydes was constrained to be planar. Solute partial atomic charges were obtained using the CHELPG[13] procedure at the HF/6-31 + G(d) level on these geometries. Constrained electrostatic potential (ESP)-derived atomic point charges were obtained using the RESP

procedure of Bayly et al.[14] from potentials defined in the CHELPG method.

The SIP was evaluated on a regular cubic grid with a 0.1-Å spacing. Lennard–Jones radii for H, C, N, O, and F were 0.80, 1.90, 2.10, 1.80, and 1.72 Å, respectively, with well depths of 0.17 kJ mol^{-1} for H, and 0.34 kJ mol^{-1} for C, N, O, and F. The offset (Γ) was 0.17 Å. FDPB calculations were performed using the DELPHI program,[15] with an external dielectric of 78.54 (water). Unless otherwise noted, an internal dielectric of 1 was used in these calculations with a grid extent of 30-Å and 0.25-Å resolution, which should be sufficient to provide results relatively independent of the grid specification.[16]

The van der Waals radii, r_{LJ}, for S, P, and Cl, for use with the SIP, have been determined here. These were obtained by minimizing the error in the calculated solvation energy for the neutral and ionized hydrides. There is, however, no reliable estimate available of the pK_a of either H_2Cl^+ or H_3S^+ and therefore no reliable estimate of their free energies of solvation; thus, these systems do not form part of the parameterization set. The experimental solvation energies of the ions have been redetermined using recent[17] experimental gas-phase acidities and basicities. Presented in Table I is a comparison the experimental and calculated solvation free energies. The optimized r_{LJ} for P, S, and Cl are 2.58, 2.30, and 2.20 Å, respectively. A Lennard–Jones well-depth of 0.52 kJ mol^{-1} is used for these atoms, with this value being taken from the parameter for the Ar–Ne interaction.[23] The calculated solvation energies all lie within 10 kJ mol^{-1} of the experimental values. The calculated solvation free energy of H_3S^+ differs by a massive 68 kJ mol^{-1} from the experimental value[22] obtained using the unreliable pK_a of −8. Using the calculated solvation free energy (−301.5 kJ mol^{-1}), the predicted pK_a of H_3S^+ is −19.9. In comparison, the pK_a predicted using the AM1-SM2 approach[24] is −16.6.

Solvation Energies of Amino Acid Side Chains

Table II presents the calculated solvation energies of 23 of the amino acid side chains, both neutral and charged. Table III presents a comparison of solvation free energies produced by the FDPB method using radii and charges from the SIP, and the PARSE parameters, as well as from the semiempirical molecular orbital (LSQM) approach (presented as differences from experiment, $\Delta\Delta G_{solv}$). For the neutral systems, the FDPB method using the PARSE parameters performs best; however, one should keep in mind that the parameters (a total of 76 in all) were fit to reproduce these values. The LSQM and SIP show comparable performance, although the SIP errors ex-

TABLE I.
Experimental and Calculated Free Energies of Solvation (kJ mol^{-1}).[a]

Molecule	$\Delta G_{acid/base}$[b]	pK_a[d]	ΔG_{expt}	$\Delta G_{nonelec}$	ΔG_{elec}	ΔG_{calc}
PH_3			2.5[c]	9.5	−1.0	8.5
PH_2^-	1520.0	27[c]	−278.1	8.9	−289.5	−280.6
PH_4^+	−750.6	−14[c]	−253.0	9.7	−261.8	−252.0
H_2S			−2.9[c]	8.6	−8.0	0.6
SH^-	1443.0	7	−320.3	8.1	−329.3	−321.2
H_3S^+	−673.8	−8[e]	−369.4	8.9	−310.4	−301.5
HCl			−2.1[c]	8.4	−12.7	−4.3
Cl^-	1372.8	−7	−328.9	7.9	−341.2	−333.3

[a] Experimental free energies of solvation of ions are determined using the following expressions:

$$\Delta G^{o}_{solv}(X^-) = -\Delta G_{acid}(HX) - \Delta G^{o}_{solv}(H^+) + \Delta G^{o}_{solv}(HX) + 5.69\ pK_a\ [\text{anions}, X^-]$$

$$\Delta G^{o}_{solv}(BH^+) = -\Delta G_{base}(BH^+) + \Delta G^{o}_{solv}(H^+) + \Delta G^{o}_{solv}(B) - 5.69\ pK_a\ [\text{cations}, BH^+]$$

where ΔG_{acid} is the gas phase acidity of HX and ΔG_{base} the gas phase basicity of B, and ΔG^{o}_{solv} the solvation free energies of the proton or parent base, HX or B. The solvation energy of the proton is taken as −1085.8 kJ mol^{-1}. Units are kilojoules per mole. [b] From ref. 17. [c] From refs. 18 and 19, with standard state corrections.[20] [d] From ref. 21, unless otherwise noted. [e] From ref. 22. The usually quoted value of −14 for PH_3 is retained. The value of −8 for H_2S is an estimate only.

TABLE II.
Calculated Amino Acid Side Chain Solvation Energies (kJ mol^{-1}).

Molecule	Amino acid	$\Delta G_{nonelec}$	ΔG_{elec}	ΔG_{solv}	ΔG_{expt}
Neutral systems					
Methane	ala	7.5	−1.1	6.4	8.4
N-propylguanidine	arg	12.0	−54.7	−42.7	−45.7
Acetamide	asn	9.7	−56.1	−47.4	−40.6
Acetic acid	asp	9.3	−53.3	−44.0	−28.0
Methylthiol	cys	9.1	−10.0	−0.9	−5.2
Propionamide	gln	10.3	−53.1	−42.8	−39.4
Propionic acid	glu	10.1	−49.2	−39.1	−27.1
Methylimidazole	his	10.3	−42.5	−32.2	−42.9
Butane	ile	9.8	−2.5	8.3	8.7
Isobutane	leu	9.8	−2.5	7.3	9.7
N-butylamine	lys	10.7	−19.2	−8.5	−17.9
Methylethylsulfide	met	10.4	−10.8	−0.4	−6.2
Toluene	phe	10.7	−12.9	−2.1	−3.7
Methanol	ser	8.1	−33.8	−25.7	−21.3
Ethanol	thr	8.9	−35.3	−26.4	−21.1
Methylindole	trp	12.0	−34.4	−22.4	−24.7
p-Cresol	tyr	11.2	−44.6	−33.4	−25.6
Propane	val	9.0	−1.1	7.9	8.2
Ionized systems					
N-propylguanidinium	arg$^+$	12.3	−258.6	−246.3	—[a]
N-butylammonium	lys$^+$	11.0	−285.9	−274.9	−277.8[b]
Methylimidazolium	met$^+$	10.8	−267.1	−256.3	−247.6[b]
Acetate ion	asp$^-$	8.8	−352.5	−343.7	−345.1[b]
Propionate ion	glu$^-$	9.5	−345.6	−336.1	−337.6[b]

[a] Experimental values unavailable. [b] Experimental solvation free energies were re-evaluated using recent[17] experimental gas phase acidities and basicities (see footnote to Table I). Solvation free energies for neutral systems taken from ref 25.

ceed 10 kJ mol^{-1} on three occasions, whereas the largest error for LSQM is 8.0 kJ mol^{-1}. Errors for the SIP method are largest for the ionizable systems (acetic acid, propionic acid, methyl imidazole, and *N*-butylamine); the largest error for the nonionizable groups is just −7.8 kJ mol^{-1}. The PARSE solvation free energies evaluated on the MP2 geometries (used in the SIP calculations) can differ from the original energies[6] by as much as 3 kJ mol^{-1}. Thus, agreement with experiment beyond this level is not significant.

The experimental solvation free energies of the ionized systems derived here are significantly different from the values used in the derivation of the PARSE parameters and the LSQM radii. This has arisen from recent significant readjustments of the gas-phase basicity and acidity scales.[17] For the ionized side chains, solvation free energies derived using the SIP parameters perform best, because both PARSE and LSQM have been fit to older experimental data, reparameterization of these methods to the newer solvation data would dramatically improve their performance. In both PARSE and LSQM, specific parameters are required for the atomic radii of the charged species; their performance with the neutral systems remains unaltered. It is worth noting that, from all the amino acid side chain molecules, only methane and methanol (whose solvation free energy is well known) form part of the data set used to define parameters of the SIP potential (unlike the PARSE set and the LSQM method, which utilized experimental information for all the amino acids).

Solvation Parameters for Amino Acids

Electrostatics play a dominant role in many biophysical processes. Solvation effects are believed to account for a large part of the forces that result in the folding of polypeptides and proteins, as well as their association with one another and

TABLE III.
Comparison of Calculated–Experimental Solvation Energies for Amino Acid Side Chains ($\Delta\Delta G_{solv}$, kJ mol^{-1}).

Molecule	Amino acid	SIP	PARSE[a]	LSQM[b]
Neutral systems				
Methane	ala	−2.0	c	−1.9
N-propylguanidine	arg	3.0	0.1	2.6
Acetamide	asn	−6.8	−0.2	−3.5
Acetic acid	asp	−16.0	0.3	−5.9
Methylthiol	cys	4.3	−0.5	−1.6
Propionamide	gln	−3.4	0.0	1.0
Propionic acid	glu	−12.0	0.3	−2.8
Methylimidazole	his	10.7	0.1	−2.3
Butane	ile	−0.4	—[c]	0.1
Isobutane	leu	−2.4	—[c]	−1.0
N-butylamine	lys	9.4	−0.3	5.8
Methylethylsulfide	met	5.8	0.1	2.6
Toluene	phe	1.6	0.0	−1.2
Methanol	ser	−4.4	−1.5	5.6
Ethanol	thr	−5.3	−0.5	8.0
Methylindole	trp	2.3	0.3	−1.9
p-Cresol	tyr	−7.8	0.1	6.9
Propane	val	−0.3	—[c]	0.0
Ionized systems				
N-propylguanidinium	arg^+	—[d]	—[d]	—[d]
N-butylammonium	lys^+	2.9	−12.7	−6.1
Methylimidazolium	met^+	−8.7	−21.7	—[e]
Acetate ion	asp^-	1.4	8.5	5.2
Propionate ion	glu^-	1.5	6.2	9.2

[a] Ref. 6. [b] Ref. 9. [c] PARSE charges on carbon and hydrogen are zero. [d] Experimental values unavaiable. [e] Value unavailable.

with smaller molecules. An accurate description of the features of these systems that describe their electrostatic potential is therefore of fundamental importance. The PARSE parameter set provides both charge and radii to enable assignment of the backbone and side chain groups of the common amino acids. The PARSE set includes only nine different radii (which includes nonpolar hydrogens set to zero). More recently, a new set of atomic radii have been developed by relating the continuum dielectric boundary to the macroscopic solvent charge distribution (SCD).[26] In that work, a total of 27 different radii were determined for amino acids. As an alternative to these, the radii can be obtained from the SIP. These parameters differ from PARSE and SCD in that all atoms, including hydrogens, for each of the amino acids, have a unique charge and radius.

The radii and charges of each of the common naturally occurring amino acids were modeled as the aminoacetaldehyde derivatives (as the simplest polypeptide building block). Chemically equivalent atoms (e.g., hydrogens of methyl groups) were constrained to have equivalent radii and partial atomic charges. Listed in Table IV are the MP2/6-31G(d)-optimized geometries, SIP-derived atomic radii, and CHELPG [RESP, HF/6-31 + G(d)] charges for the aminoacetaldehydes.

Apart from the quantity of radii and charge data, perhaps the most obvious difference between SIP-derived parameters and other parameter sets is the charges on nonpolar carbons. Generally these, and the associated hydrogens, are constrained to have zero charge. In all cases, the SIP radii of the hydrogens attached to carbon extend beyond the carbon radius, but by less than 0.1 Å. In addition, the charges on these hydrogen atoms never exceed 0.1. Simplification of the parameters could involve setting the radius of these hydrogen atoms to zero, collapsing their charge to the carbon, and increasing the carbon radius by $\sim$ 0.1 Å. Note that no generalization can be made about the carbon atoms; the radii differ by 0.1 Å and the charges vary between −0.4 and +0.9. Even when hydro-

TABLE IV.
Geometries, Atomic Radii, and Charges for Amino Acids.

	X	Y	Z	R	Q
Alanine					
N	−0.334	1.396	−0.022	1.97	−1.322
HN1	0.558	1.724	−0.352	1.04	0.441
HN2	−1.120	2.020	−0.015	1.12	0.476
CA	−0.447	0.052	0.424	1.86	0.826
HA	−0.633	−0.033	1.514	0.79	−0.143
C	0.870	−0.652	0.205	1.81	0.415
HC	0.886	−1.718	0.510	0.88	−0.028
O	1.864	−0.120	−0.269	1.58	−0.521
CB	−1.552	−0.738	−0.289	1.84	−0.358
HB1	−1.365	−0.744	−1.365	0.88	0.071
HB2	−2.519	−0.261	−0.111	0.88	0.071
HB3	−1.613	−1.768	0.077	0.88	0.071
Arginine(+)					
N	−3.830	1.166	−0.311	2.05	−1.391
HN1	−3.850	2.061	−0.767	1.09	0.496
HN2	−4.568	0.923	0.329	1.08	0.485
CA	−2.792	0.232	−0.544	1.87	0.840
HA	−2.655	0.003	−1.618	0.83	−0.119
C	−3.157	−1.079	0.122	1.82	0.387
HC	−2.501	−1.941	−0.107	0.83	−0.014
O	−4.111	−1.209	0.873	1.63	−0.490
CB	−1.427	0.719	0.004	1.88	−0.259
HB1	−1.525	0.837	1.090	0.81	0.060
HB2	−1.263	1.722	−0.409	0.81	0.060
CG	−0.232	−0.171	−0.340	1.86	−0.113
HG1	−0.206	−0.355	−1.421	0.90	0.044
HG2	−0.315	−1.141	0.162	0.90	0.044
CD	1.071	0.488	0.088	1.91	0.401
HD1	1.054	0.672	1.168	0.87	0.003
HD2	1.195	1.444	−0.434	0.87	0.003
NE	2.206	−0.388	−0.233	2.03	−0.824
HE	1.996	−1.265	−0.699	1.02	0.397
CZ	3.478	−0.111	0.036	2.04	1.239
NH1	4.448	−0.973	−0.298	2.09	−1.195
HH11	4.759	1.267	0.865	1.08	0.491
HH12	3.093	1.706	0.910	1.10	0.462
NH2	3.800	1.037	0.641	2.08	−1.055
HH21	4.240	−1.853	−0.751	1.08	0.522
HH22	5.420	−0.783	−0.098	1.08	0.526
Arginine					
N	−3.783	1.172	−0.289	1.95	−1.475
HN1	−4.596	0.887	0.230	1.05	0.491
HN2	−3.720	2.111	−0.639	1.12	0.501
CA	−2.743	0.228	−0.509	1.81	0.976
HA	−2.610	−0.032	−1.579	0.79	−0.160
C	−3.120	−1.068	0.167	1.80	0.350
HC	−2.406	−1.903	0.038	0.86	0.006
O	−4.142	−1.223	0.824	1.57	−0.527
CB	−1.373	0.711	0.011	1.81	−0.222
HB1	−1.443	0.827	1.099	0.80	0.024
HB2	−1.221	1.719	−0.397	0.80	0.024
CG	−0.186	−0.167	−0.375	1.77	−0.281
HG1	−0.186	−0.355	−1.455	0.76	0.053
Asparagine					
N	0.907	1.271	−0.563	1.92	−1.237
HN1	1.714	1.311	−1.161	1.05	0.441
HN2	0.184	1.965	−0.621	1.04	0.455
CA	0.795	0.213	0.374	1.81	0.587
HA	0.619	0.582	1.400	0.75	−0.058
C	2.080	−0.573	0.417	1.81	0.504
HC	2.111	−1.385	1.172	0.90	−0.037
O	3.039	−0.361	−0.311	1.58	−0.540
CB	−0.343	−0.770	0.056	1.84	−0.364
HB1	−0.173	−1.200	−0.938	0.85	0.080
HB2	−0.379	−1.599	0.778	0.85	0.080
CG	−1.679	−0.048	0.088	1.77	0.943
OD1	−1.823	1.066	0.589	1.54	−0.653
ND2	−2.727	−0.757	−0.428	2.04	−1.084
HD21	−2.575	−1.567	−1.011	1.07	0.440
HD22	−3.612	−0.274	−0.513	1.06	0.441
Aspartate					
N	1.843	1.353	0.283	1.85	−1.535
HN1	1.690	2.329	0.454	1.03	0.485
HN2	2.754	1.024	0.014	0.95	0.496
CA	0.778	0.401	0.414	1.64	1.025
HA	0.490	0.194	1.460	0.63	−0.169
C	1.253	−0.924	−0.115	1.64	0.368
HC	0.486	−1.713	−0.099	0.75	0.036
O	2.395	−1.128	−0.531	1.54	−0.621
CB	−0.513	0.808	−0.309	1.63	−0.348
HB1	−0.756	1.843	−0.030	0.63	0.037
HB2	−0.339	0.828	−1.393	0.63	0.037
CG	−1.758	−0.081	0.007	1.63	0.966
OD1	−2.824	0.296	−0.541	1.51	−0.889
OD2	−1.546	−1.068	0.777	1.51	−0.889
Aspartic acid					
N	1.909	−1.311	0.150	2.00	−1.373
HN1	2.768	−0.838	0.377	1.03	0.472
HN2	1.858	−2.311	0.220	1.12	0.491
CA	0.798	−0.535	−0.263	1.85	0.822
HA	0.415	−0.822	−1.259	0.81	−0.107
C	1.227	0.914	−0.375	1.79	0.336
HC	0.445	1.624	−0.700	0.86	0.017
O	2.361	1.300	−0.128	1.58	−0.493
CB	−0.399	−0.618	0.726	1.83	−0.295
HB1	−0.628	−1.676	0.882	0.85	0.088
HB2	−0.116	−0.170	1.681	0.85	0.088
CG	−1.622	0.077	0.200	1.86	0.810
OD1	−2.024	1.179	0.529	1.58	−0.634
OD2	−2.230	−0.669	−0.763	1.74	−0.669
HD2	−2.989	−0.132	−1.074	0.92	0.448
Cystine					
N	−1.616	1.339	−0.042	1.98	−1.468
HN1	−2.455	0.955	0.359	1.03	0.485
HN2	−1.540	2.331	−0.175	1.13	0.514
CA	−0.566	0.456	−0.404	1.85	0.904

(Continued)

TABLE IV.
(Continued)

	X	Y	Z	R	Q
Arginine					
HG2	−0.236	−1.146	0.113	0.76	0.053
CD	1.146	0.468	−0.002	1.79	0.893
HD1	1.151	0.666	1.084	0.79	−0.146
HD2	1.230	1.447	−0.502	0.79	−0.146
NE	2.231	−0.414	−0.401	1.74	−1.040
CZ	3.414	−0.130	0.013	1.84	1.104
NH1	3.830	1.008	0.702	2.01	−1.027
HH11	4.160	−1.757	−0.844	1.02	0.415
HH12	5.297	−0.556	−0.589	1.08	0.405
NH2	4.470	−1.017	−0.221	1.99	−1.056
HH21	4.564	0.825	1.378	1.08	0.419
HH22	3.080	1.560	1.100	1.08	0.364
Glutamate					
N	1.764	1.244	−0.812	1.87	−1.243
HN1	2.511	0.898	−1.388	0.95	0.399
HN2	1.205	2.016	−1.126	1.05	0.458
CA	1.507	0.634	0.448	1.68	0.731
HA	1.964	1.185	1.304	0.60	−0.198
C	2.168	−0.714	0.500	1.72	0.479
HC	1.925	−1.307	1.407	0.75	−0.065
O	2.919	−1.167	−0.357	1.56	−0.578
CB	0.008	0.494	0.758	1.65	0.026
HB1	−0.468	1.479	0.700	0.69	−0.036
HB2	−0.100	0.165	1.801	0.69	−0.036
CG	−0.766	−0.425	−0.172	1.60	−0.027
HG1	−0.579	−1.484	0.047	0.63	−0.040
HG2	−0.439	−0.246	−1.204	0.63	−0.040
CD	−2.309	−0.162	−0.106	1.67	0.942
OE1	−3.036	−1.172	−0.298	1.52	−0.886
OE2	−2.634	1.043	0.102	1.52	−0.886
Glutamic acid					
N	1.865	1.336	−0.604	2.00	−1.227
HN1	2.404	0.899	−1.333	1.05	0.419
HN2	1.560	2.286	−0.717	1.10	0.454
CA	1.558	0.588	0.563	1.88	0.702
HA	1.997	1.022	1.483	0.81	−0.127
C	2.170	−0.787	0.444	1.82	0.405
HC	2.015	−1.445	1.322	0.89	−0.017
0	2.794	−1.178	−0.532	1.58	−0.511
CB	0.046	0.431	0.828	1.82	−0.012
HB1	−0.381	1.427	0.980	0.86	−0.017
HB2	−0.108	−0.128	1.761	0.86	−0.017
CG	−0.667	−0.260	−0.334	1.81	−0.063
HG1	−0.349	−1.301	−0.441	0.81	0.029
HG2	−0.419	0.273	−1.257	0.81	0.029
CD	−2.157	−0.270	−0.159	1.84	0.802
OE1	−2.867	−1.254	−0.056	1.58	−0.647
OE2	−2.658	0.997	−0.115	1.72	−0.638
HE2	−3.624	0.890	0.007	0.93	0.436
Glutamine					
N	2.403	1.379	−0.044	1.97	−1.434
HN1	3.286	1.094	−0.433	1.04	0.494
HN2	2.173	2.355	0.003	1.12	0.498
CA	1.505	0.387	0.427	1.85	0.810
Cystine					
HA	−0.270	0.537	−1.465	0.83	−0.121
C	−1.042	−0.965	−0.213	1.79	0.353
HC	−0.335	−1.754	−0.527	0.87	−0.023
O	−2.135	−1.255	0.255	1.58	−0.499
CB	0.706	0.703	0.438	1.83	−0.175
HB1	0.973	1.757	0.333	0.86	0.094
HB2	0.485	0.532	1.494	0.86	0.094
SG	2.191	−0.217	−0.077	2.23	−0.340
HG	1.897	−1.374	0.533	1.03	0.184
Histidine(+)					
N	−1.418	1.253	−0.606	2.04	−1.199
HN1	−2.120	1.785	−0.114	1.07	0.438
HN2	−0.734	1.738	−1.157	1.04	0.428
CA	−1.438	−0.156	−0.488	1.89	0.700
HA	−1.187	−0.645	−1.442	0.87	−0.096
C	−2.835	−0.613	−0.089	1.86	0.411
HC	−3.037	−1.698	−0.168	0.90	0.002
O	−3.688	0.169	0.296	1.67	−0.472
CB	−0.476	−0.720	0.611	1.85	−0.307
HB1	−0.799	−0.320	1.578	0.89	0.090
HB2	−0.532	−1.814	0.671	0.89	0.090
CG	0.923	−0.293	0.365	1.92	0.272
ND1	1.884	−1.089	−0.227	2.05	−0.322
HD1	1.750	−2.063	−0.494	1.04	0.376
CD2	1.536	0.924	0.565	1.89	−0.122
HD2	1.155	1.832	1.006	0.90	0.214
CE1	3.037	−0.424	−0.378	1.91	0.178
HE1	3.956	−0.805	−0.799	0.91	0.207
NE2	2.826	0.806	0.106	2.07	−0.289
HE2	3.531	1.539	0.131	1.06	0.401
Histidine (δH)					
N	2.823	−0.992	0.100	1.99	−1.372
HN1	3.547	−0.311	0.256	1.03	0.465
HN2	3.025	−1.967	0.227	1.12	0.483
CA	1.538	−0.542	−0.296	1.83	0.797
HA	1.202	−0.963	−1.262	0.81	−0.136
C	1.578	0.954	−0.494	1.79	0.450
HC	0.653	1.408	−0.894	0.90	−0.040
O	2.557	1.645	−0.238	1.57	−0.540
CB	0.430	−0.878	0.742	1.79	−0.150
HB1	0.450	−1.962	0.898	0.84	0.052
HB2	0.713	−0.423	1.699	0.84	0.052
CG	−0.938	−0.458	0.326	1.73	−0.024
ND1	−1.545	0.698	0.772	2.02	−0.341
HD1	−1.175	1.325	1.476	1.04	0.314
CD2	−1.839	−0.981	−0.582	1.75	0.062
HD2	−1.737	−1.892	−1.158	0.87	0.119
CE1	−2.753	0.810	0.145	1.82	0.310
HE1	−3.434	1.629	0.337	0.89	0.079
NE2	−2.962	−0.196	−0.690	1.83	−0.579

(Continued)

TABLE IV.
(Continued)

	X	Y	Z	R	Q
Glutamine					
HA	1.351	0.420	1.527	0.79	−0.119
C	2.097	−0.974	0.158	1.79	0.371
HC	1.500	−1.835	0.516	0.85	0.004
O	3.163	−1.157	−0.415	1.58	−0.524
CB	0.111	0.499	−0.222	1.79	−0.051
HB1	0.216	0.337	−1.301	0.80	0.023
HB2	−0.238	1.527	−0.091	0.80	0.023
CG	−0.947	−0.429	0.367	1.85	−0.238
HG1	−0.803	−1.467	0.048	0.85	0.032
HG2	−0.882	−0.420	1.464	0.85	0.032
CD	−2.348	0.038	0.001	1.76	0.976
OE1	−2.641	1.226	−0.120	1.55	−0.684
NE2	−3.261	−0.961	−0.183	2.06	−1.103
HE22	−3.057	−1.920	0.052	1.06	0.441
HE21	−4.229	−0.692	−0.298	1.06	0.447
Glycine					
N	−1.315	−0.547	0.000	1.99	−1.247
HN1	−0.774	−1.394	0.000	1.03	0.425
HN2	−2.316	−0.597	0.000	1.18	0.456
CA	−0.625	0.692	0.000	1.87	0.688
HA1	−0.836	1.330	0.879	0.84	−0.103
HA2	−0.836	1.330	−0.879	0.84	−0.103
C	0.863	0.463	0.000	1.83	0.407
HC	1.482	1.381	0.000	0.89	−0.013
O	1.382	−0.644	0.000	1.58	−0.511
Isoleucine					
N	−1.102	−1.604	0.196	1.95	−1.293
HN1	−1.857	−1.787	−0.443	1.04	0.458
HN2	−0.594	−2.372	0.593	1.11	0.469
CA	−0.792	−0.250	0.507	1.85	0.491
HA	−0.689	−0.095	1.598	0.80	−0.069
C	−1.960	0.606	0.067	1.80	0.509
HC	−1.975	1.647	0.436	0.86	−0.033
O	−2.857	0.197	−0.659	1.58	−0.549
CB	0.519	0.288	−0.142	1.84	0.138
HB	0.375	0.221	−1.231	0.75	−0.030
CG1	1.685	−0.627	0.235	1.80	0.275
HG11	1.399	−1.656	0.001	0.78	−0.069
HG12	1.837	−0.576	1.323	0.78	−0.069
CG2	0.796	1.740	0.248	1.81	−0.345
HG21	0.830	1.845	1.339	0.87	0.071
HG22	0.040	2.429	−0.135	0.87	0.071
HG23	1.758	2.077	−0.145	0.87	0.071
CD1	2.994	−0.315	−0.485	1.81	−0.257
HD11	3.404	0.656	−0.200	0.86	0.053
HD12	3.750	−1.070	−0.251	0.86	0.053
HD13	2.847	−0.315	−1.569	0.86	0.053
Leucine					
N	1.176	1.127	−0.874	1.94	−1.281
HN1	2.068	1.029	−1.329	1.05	0.447
HN2	0.509	1.785	−1.232	1.08	0.457
CA	0.903	0.322	0.265	1.84	0.656
HA	0.719	0.918	1.178	0.79	−0.111
Histidine (εH)					
N	2.706	−1.185	−0.123	1.93	−1.391
HN1	3.544	−0.654	0.041	1.02	0.467
HN2	2.745	−2.188	−0.132	1.12	0.475
CA	1.491	−0.482	−0.338	1.76	0.940
HA	1.063	−0.641	−1.348	0.66	−0.195
C	1.772	1.003	−0.248	1.73	0.439
HC	0.899	1.663	−0.372	0.81	−0.021
O	2.893	1.458	−0.049	1.56	−0.562
CB	0.383	−0.867	0.677	1.71	−0.399
HB1	0.356	−1.962	0.726	0.82	0.086
HB2	0.679	−0.512	1.670	0.82	0.086
CG	−0.974	−0.361	0.319	1.71	0.395
ND1	−1.316	0.970	0.451	1.79	−0.587
CD2	−2.037	−1.071	−0.201	1.83	−0.294
HD2	−2.164	−2.118	−0.439	0.92	0.194
CE1	−2.565	1.061	0.021	1.84	0.267
HE1	−3.161	1.963	−0.015	0.90	0.100
NE2	−3.041	−0.152	−0.382	2.05	−0.347
HE2	−3.968	−0.343	−0.740	1.09	0.346
Lysine					
N	2.903	1.292	0.005	1.95	−1.461
HN1	3.749	0.924	−0.395	1.04	0.498
HN2	2.788	2.285	0.094	1.12	0.499
CA	1.893	0.388	0.432	1.84	0.834
HA	1.729	0.403	1.529	0.80	−0.115
C	2.344	−1.019	0.121	1.79	0.363
HC	1.656	−1.824	0.439	0.85	−0.003
O	3.395	−1.290	−0.444	1.58	−0.520
CB	0.526	0.662	−0.231	1.81	−0.044
HB1	0.629	0.495	−1.310	0.80	0.003
HB2	0.323	1.733	−0.105	0.80	0.003
CG	−0.649	−0.132	0.334	1.82	−0.153
HG1	−0.527	−1.202	0.127	0.84	−0.006
HG2	−0.667	−0.028	1.428	0.84	−0.006
CD	−1.985	0.335	−0.235	1.79	0.142
HD1	−2.146	1.394	−0.007	0.81	−0.015
HD2	−1.963	0.250	−1.330	0.81	−0.015
CE	−3.169	−0.455	0.303	1.84	0.343
HE1	−3.012	−1.528	0.096	0.84	−0.027
HE2	−3.208	−0.347	1.392	0.84	−0.027
NZ	−4.419	0.080	−0.242	1.98	−0.970
HZ1	−5.212	−0.421	0.157	1.11	0.339
HZ2	−4.451	−0.099	−1.245	1.11	0.339
Methionine					
N	1.818	1.457	−0.221	1.97	−1.218
HN1	2.616	1.514	−0.832	1.03	0.428
HN2	1.216	2.253	−0.122	1.10	0.433

(Continued)

TABLE IV.
(Continued)

	X	Y	Z	R	Q
Leucine					
C	2.124	−0.506	0.582	1.80	0.499
HC	2.032	−1.148	1.481	0.87	−0.040
O	3.154	−0.494	−0.077	1.58	−0.547
CB	−0.282	−0.647	0.089	1.83	−0.302
HB1	−0.419	−1.229	1.013	0.81	0.031
HB2	−0.024	−1.352	−0.711	0.81	0.031
CG	−1.603	0.039	−0.262	1.81	0.560
HG	−1.453	0.578	−1.207	0.78	−0.089
CD1	−2.690	−1.008	−0.491	1.80	−0.397
HD11	−2.397	−1.719	−1.269	0.86	0.080
HD12	−3.630	−0.539	−0.797	0.86	0.080
HD13	−2.879	−1.572	0.429	0.86	0.080
CD2	−2.038	1.039	0.805	1.80	−0.397
HD21	−1.321	1.856	0.927	0.86	0.080
HD22	−3.004	1.483	0.547	0.86	0.080
HD23	−2.149	0.541	1.775	0.86	0.080
Lysine(+)					
N	2.986	1.274	−0.044	2.05	−1.386
HN1	3.736	0.882	−0.589	1.08	0.490
HN2	2.972	2.262	0.134	1.10	0.496
CA	1.977	0.411	0.450	1.88	0.748
HA	1.844	0.491	1.546	0.82	−0.091
C	2.388	−1.022	0.182	1.83	0.402
HC	1.763	−1.805	0.656	0.84	−0.010
O	3.345	−1.329	−0.511	1.62	−0.493
CB	0.599	0.678	−0.207	1.83	−0.164
HB1	0.700	0.489	−1.282	0.83	0.045
HB2	0.399	1.750	−0.093	0.83	0.045
CG	−0.569	−0.118	0.375	1.86	−0.029
HG1	−0.441	−1.189	0.182	0.89	0.000
HG2	−0.597	0.009	1.465	0.89	0.000
CD	−1.901	0.338	−0.225	1.91	0.023
HD1	−2.051	1.403	−0.009	0.87	0.024
HD2	−1.861	0.226	−1.316	0.87	0.024
CE	−3.054	−0.471	0.334	1.91	0.161
HE1	−2.964	−1.535	0.104	0.93	0.076
HE2	−3.163	−0.349	1.413	0.93	0.076
NZ	−4.367	−0.016	−0.280	2.09	−0.542
HZ1	−4.540	0.979	−0.084	1.07	0.369
HZ2	−4.353	−0.126	−1.302	1.07	0.369
HZ3	−5.172	−0.548	0.078	1.07	0.369
Proline					
N	0.319	−0.974	0.641	1.96	−0.680
HN	−0.022	−1.833	1.039	1.06	0.317
CA	−0.377	0.265	0.793	1.85	0.274
HA	−0.615	0.516	1.843	0.81	−0.049
C	−1.686	0.335	0.023	1.81	0.575
HC	−2.261	1.276	0.179	0.85	−0.063
O	−2.109	−0.540	−0.716	1.58	−0.567
CB	0.624	1.287	0.219	1.83	−0.084
HB1	0.138	2.176	−0.197	0.85	0.027
HB2	1.302	1.615	1.013	0.85	0.027
CG	1.383	0.461	−0.821	1.81	−0.023
HG1	2.337	0.911	−1.108	0.85	0.020
HG2	0.772	0.340	−1.721	0.85	0.020
Methionine					
CA	1.585	0.241	0.474	1.87	0.628
HA	1.499	0.393	1.568	0.79	−0.112
C	2.775	−0.670	0.282	1.81	0.456
HC	2.729	−1.629	0.835	0.88	−0.031
O	3.735	−0.400	−0.426	1.58	−0.530
CB	0.329	−0.535	0.024	1.80	−0.006
HB1	0.260	−1.481	0.576	0.80	0.007
HB2	0.431	−0.769	−1.041	0.80	0.007
CG	−0.935	0.280	0.251	1.83	−0.116
HG1	−0.889	1.198	−0.343	0.85	0.086
HG2	−1.020	0.554	1.309	0.85	0.086
SD	−2.390	−0.677	−0.246	2.14	−0.259
CE	−3.656	0.573	0.064	1.86	−0.023
HE1	−4.622	0.135	−0.194	0.88	0.055
HE2	−3.677	0.865	1.117	0.88	0.055
HE3	−3.499	1.458	−0.556	0.88	0.055
Phenylalanine					
N	−1.580	1.054	−0.878	1.93	−1.247
HN1	−2.419	1.607	−0.925	1.06	0.433
HN2	−0.708	1.434	−1.200	1.07	0.423
CA	−1.657	−0.259	−0.350	1.82	0.810
HA	−1.397	−1.044	−1.089	0.75	−0.169
C	−3.076	−0.549	0.069	1.81	0.439
HC	−3.245	−1.566	0.477	0.89	−0.025
O	−3.996	0.252	−0.020	1.58	−0.542
CB	−0.734	−0.485	0.874	1.79	−0.273
HB1	−0.866	−1.507	1.255	0.83	0.040
HB2	−1.038	0.214	1.661	0.83	0.040
CG	0.702	−0.257	0.497	1.70	0.196
CD1	1.290	1.004	0.649	1.77	−0.185
HD1	0.714	1.808	1.103	0.87	0.120
CD2	1.455	−1.284	−0.084	1.77	−0.185
HD2	1.012	−2.272	−0.203	0.87	0.120
CE1	2.605	1.230	0.242	1.78	−0.091
HE1	3.050	2.213	0.374	0.90	0.107
HE2	3.343	−1.873	−0.933	0.90	0.107
CE2	2.770	−1.062	−0.491	1.78	−0.091
CZ	3.347	0.199	−0.334	1.77	−0.140
HZ	4.373	0.373	−0.648	0.92	0.114
Tryptophan					
N	2.404	0.670	1.203	1.90	−1.345
HN1	3.339	0.871	1.515	1.04	0.458
HN2	1.626	1.194	1.560	1.10	0.445
CA	2.205	−0.364	0.255	1.76	0.853
HA	1.635	−1.227	0.657	0.76	−0.171
C	3.537	−0.922	−0.173	1.79	0.393
HC	3.483	−1.746	−0.914	0.88	−0.019
O	4.617	−0.523	0.241	1.57	−0.537
CB	1.444	0.120	−1.004	1.74	0.034
HB1	2.035	0.918	−1.469	0.78	−0.037
HB2	1.356	−0.700	−1.731	0.78	−0.037
CG	0.092	0.629	−0.639	1.62	−0.147
CD1	−0.272	1.946	−0.452	1.80	−0.018

(Continued)

TABLE IV.
(Continued)

	X	Y	Z	R	Q
Proline					
CD	1.541	−0.891	−0.125	1.81	0.244
HD1	1.628	−1.713	−0.845	0.78	−0.019
HD2	2.448	−0.901	0.503	0.78	−0.019
Serine					
N	0.205	1.485	−0.270	2.03	−1.087
HN1	0.983	1.424	−0.907	1.05	0.400
HN2	−0.408	2.280	−0.310	1.10	0.407
CA	0.008	0.431	0.669	1.88	0.480
HA	0.124	0.748	1.724	0.84	−0.090
C	1.054	−0.633	0.423	1.81	0.471
HC	0.976	−1.527	1.072	0.87	−0.008
O	1.932	−0.540	−0.421	1.58	−0.527
CB	−1.368	−0.224	0.506	1.83	0.201
HB1	−2.149	0.516	0.738	0.86	0.012
HB2	−1.489	−1.065	1.196	0.86	0.012
OG	−1.497	−0.733	−0.809	1.58	−0.650
HG	−1.116	−0.030	−1.372	1.01	0.379
Threonine					
N	0.797	1.540	0.047	1.91	−1.267
HN1	1.758	1.644	−0.230	1.06	0.450
HN2	0.175	2.327	0.063	1.05	0.464
CA	0.318	0.260	0.419	1.81	0.576
HA	0.060	0.175	1.496	0.77	−0.047
C	1.394	−0.771	0.177	1.79	0.417
HC	1.147	−1.809	0.466	0.86	−0.020
O	2.486	−0.507	−0.309	1.58	−0.518
CB	−0.972	−0.098	−0.336	1.84	0.415
HB	−0.743	−0.057	−1.411	0.77	−0.045
OG1	−1.867	0.965	0.016	1.60	−0.748
HG1	−2.641	0.900	−0.569	0.96	0.446
CG2	−1.580	−1.440	0.038	1.85	−0.399
HG21	−0.963	−2.281	−0.290	0.89	0.093
HG22	−2.560	−1.549	−0.436	0.89	0.093
HG23	−1.722	−1.500	1.121	0.89	0.093
Valine					
N	0.823	1.558	0.120	1.96	−1.248
HN1	1.697	1.624	−0.373	1.07	0.434
HN2	0.331	2.397	0.368	1.12	0.460
CA	0.323	0.267	0.445	1.83	0.513
HA	0.071	0.177	1.520	0.80	−0.077
C	1.424	−0.738	0.191	1.79	0.458
HC	1.245	−1.764	0.560	0.85	−0.035
O	2.469	−0.464	−0.387	1.58	−0.528
CB	−0.953	−0.147	−0.340	1.82	0.285
HB	−0.673	−0.170	−1.402	0.76	−0.055
CG1	−1.482	−1.519	0.078	1.82	−0.253
HG11	−0.805	−2.335	−0.186	0.86	0.050
HG12	−2.439	−1.721	−0.412	0.86	0.050
HG13	−1.652	−1.554	1.160	0.86	0.050
CG2	−2.036	0.910	−0.152	1.82	−0.253
HG21	−1.700	1.887	−0.505	0.86	0.050
HG22	−2.936	0.642	−0.713	0.86	0.050
HG23	−2.311	0.993	0.906	0.86	0.050
Tryptophan					
HD1	0.309	2.848	−0.600	0.89	0.160
CD2	−1.058	−0.159	−0.307	1.62	0.063
NE1	−1.587	2.010	−0.043	2.00	−0.610
HE1	−2.103	2.863	0.128	1.10	0.410
CE2	−2.101	0.738	0.049	1.67	0.324
CE3	−1.323	−1.541	−0.307	1.65	−0.150
HE3	−0.550	−2.255	−0.584	1.78	0.116
CZ2	−3.376	0.299	0.421	1.73	−0.295
HZ2	−4.163	0.999	0.694	0.90	0.149
CZ3	−2.585	−1.981	0.061	1.69	−0.191
HZ3	−2.802	−3.046	0.068	0.87	0.121
CH2	−3.602	−1.071	0.418	1.73	−0.076
HH2	−4.581	−1.449	0.701	0.89	0.108
Tyrosine					
N	1.942	−1.090	−0.867	1.92	−1.246
HN1	2.761	−1.669	−0.942	1.05	0.437
HN2	1.044	−1.450	−1.137	1.11	0.407
CA	2.083	0.232	−0.377	1.82	0.797
HA	1.819	1.008	−1.125	0.76	−0.147
C	3.527	0.485	−0.024	1.80	0.408
HC	3.745	1.505	0.354	0.88	−0.018
O	4.418	−0.346	−0.133	1.58	−0.531
CB	1.219	0.516	0.876	1.79	−0.215
HB1	1.534	−0.172	1.668	0.83	0.029
HB2	1.396	1.542	1.226	0.83	0.029
CG	−0.238	0.322	0.569	1.68	0.078
CD1	−0.989	1.348	−0.024	1.76	−0.104
HD1	−0.525	2.317	−0.202	0.87	0.121
CD2	−0.863	−0.909	0.788	1.76	−0.104
HD2	−0.302	−1.719	1.249	0.87	0.121
CE1	−2.323	1.162	−0.368	1.77	−0.313
HE1	−2.909	1.958	−0.818	0.88	0.169
CE2	−2.201	−1.110	0.448	1.77	−0.313
HE2	−2.673	−2.073	0.636	0.88	0.169
CZ	−2.930	−0.074	−0.137	1.69	0.435
OH	−4.248	−0.197	−0.500	1.63	−0.589
HH	−4.546	−1.097	−0.278	0.95	0.379

gen charges are combined into the carbon, the charges vary between -0.2 and $+0.6$. The SIP carbon radii (1.74–1.86 Å) are comparable in size to the PARSE value (2.00 Å), but substantially smaller than the CB (carbon beta) radii (2.67 Å) from SCD. The largest difference occurs for the CA (carbon alpha); the SIP radii vary between 1.81 and 1.88 Å, whereas the SCD value is 2.86 Å. The SIP carbon and hydrogen radii on aspartate and glutamate are roughly 0.2 Å smaller than their neutral counterparts, whereas they are slightly larger for the cations arginine and lysine. The hydroxyl oxygen radii (Ser, Thr, Tyr) from SIP (1.58–1.63 Å) are intermediate in size between the PARSE (1.40 Å) and SCD (1.85 Å) radii. Because the SIP and PARSE charges are both derived from the electrostatic potential charges from molecular orbital calculations, they are generally rather similar.

Transferability to Polypeptides

Forming the peptide unit from the aminoacetaldehydes has been achieved by removing the aldehyde hydrogen and one of the amine hydrogens and combining their charge on the carbon and nitrogen, respectively. An alternative approach in which the charges on the hydrogens were constrained to zero in the electrostatic potential fitting procedure yielded significantly poorer solvation energies. The electrostatic component of the solvation free energy of di-glycine (intact) calculated using the SIP approach is -74.6 kJ mol^{-1}. This compares with -73.8 kJ mol^{-1} when the SIP charges and radii for the single glycine aminoacetaldehyde are used to construct the dipeptide, a difference of 0.8 kJ mol^{-1}. The difference is -2.4, 3.6, and 6.0 kJ mol^{-1} using the parameters from either source for the alanine and serine dipeptides and the disulfide-linked cystine dipeptide.

For the glycine dipeptide, the PARSE parameters predict a solvation energy of -85.5 kJ mol^{-1} (which differs substantially from the SIP-derived solvation energy of -74.6 kJ mol^{-1}), whereas for the cystine dipeptide using SIP and PARSE parameters yields solvation energies that differ by just 4.6 kJ mol^{-1}.

Binding Energies in Trypsin

Trypsin binds many derivatives of benzamidine that inhibit its function.[27] Five of these derivatives,

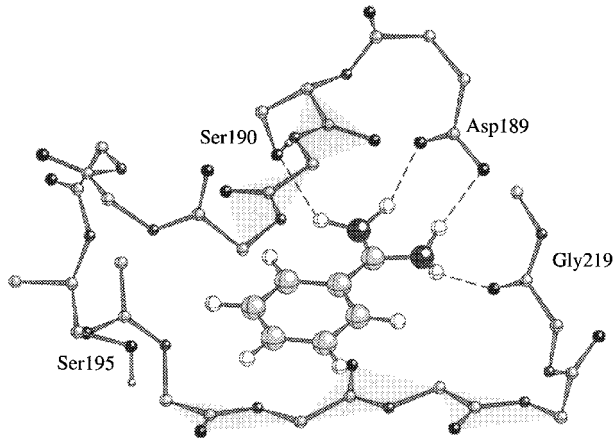

FIGURE 1. Interactions between benzamidine and the active site residues of the protein.

p-amino-, *p*-chloro-, *p*-fluoro-, *p*-hydroxy-, and *p*-methylbenzamidine, are all believed to bind in a similar manner to the parent benzamidine. The X-ray crystal structure of benzamidine bound to trypsin[28] shows that the amidinium group makes hydrogen bonds with a backbone carbonyl of the protein (Gly219), as well as the side chain groups of an aspartic acid (Asp189) and serine (Ser190). The ring of the benzyl group lies between the hydrophobic faces of two extended backbone strands. In addition to these interactions, the *p*-NH_2 and *p*-OH groups are believed to make contact with a serine residue of the protein (Ser195). Presented in Figure 1 is a diagram illustrating the interactions between benzamidine and the active-site residues of the protein.

Calculation of the binding energy of inhibitors of typsin has attracted a great deal of attention from a wide variety of perspectives.[30–34] The benzamidine inhibitors of trypsin provide a useful test for calculating the electrostatic contribution to the energy of small ligands binding to a protein. First, they contain no freely rotatable bonds (ignoring the methyl rotation in *p*-methylbenzamidine). There is, therefore, no ambiguity regarding the geometry of the compounds, either in the active site of the protein or in solution. Second (with the exception of *p*-NH_2 and *p*-OH derivatives), they all bind with the same interactions with the protein. Nonelectrostatic contributions to the relative binding energy should roughly cancel. And, finally, because the substituents in the para position remain relatively exposed to the solvent, the nonelectrostatic component of the solvation energy in free solution roughly cancels the contribution when bound to the protein.

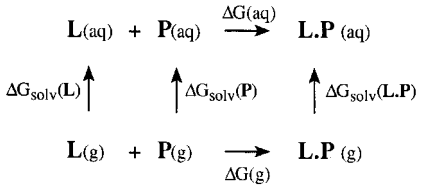

FIGURE 2. Thermodynamic cycle shows how the binding energy between a ligand, **L**, and a protein, **P**, in solution, $\Delta G(\mathrm{aq})$, is evaluated from knowledge of the solvation free energies, ΔG_{solv}, of the individual components, and the gas phase interaction energy, $\Delta G(\mathrm{g})$.

Electrostatic methods have been used previously to calculate successfully the binding energies of ligands to proteins.[35,36] The relative binding affinity of the benzamidine inhibitors of trypsin have been calculated here making use of the thermodynamic cycle shown in Figure 2. The total binding free energy between a ligand, **L**, and the protein, **P**, $\Delta G(\mathrm{aq})$, is related to the gas-phase binding energy, $\Delta G(\mathrm{g})$, and the solvation free energies for the ligand, protein, and complex, by the expression:

$$\Delta G(\mathrm{aq}) = \Delta G(\mathrm{g}) + \Delta G_{\mathrm{solv}}(\mathbf{L}) + \Delta G_{\mathrm{solv}}(\mathbf{P}) - \Delta G_{\mathrm{solv}}(\mathbf{L.P})$$

The relative binding energy for two ligands **L1** and **L2** is thus given by:

$$\Delta\Delta G_{12} = \Delta G_2(\mathrm{g}) - \Delta G_1(\mathrm{g}) + \Delta G_{\mathrm{solv}}(\mathbf{L2}) - \Delta G_{\mathrm{solv}}(\mathbf{L1}) - \Delta G_{\mathrm{solv}}(\mathbf{L1.P}) + \Delta G_{\mathrm{solv}}(\mathbf{L2.P})$$

which can be rewritten as:

$$\Delta\Delta G_{12} = \Delta\Delta G_{\mathrm{inter}} + \Delta\Delta G_{\mathrm{solv}}(\mathbf{L}) - \Delta\Delta G_{\mathrm{solv}}(\mathbf{L.P})$$

where $\Delta\Delta G_{\mathrm{inter}}$ is the difference in the free energy of interaction between the two ligands and the protein, $\Delta\Delta G_{\mathrm{solv}}(\mathbf{L})$ is the difference in the solvation energy of the two ligands, and $\Delta\Delta G_{\mathrm{solv}}(\mathbf{L.P})$ is the difference in solvation energy of the two ligand–protein complexes.

The X-ray crystallographic coordinates of the non-hydrogen atoms of the protein complexed with benzamidine were taken from the Brookhaven Protein Data Bank (structure 3ptb).[28] Hydrogen atoms were added to satisfy unfilled valencies; in doing so, ionizable residues lying more than 10 Å from the inhibitor were maintained with neutral charge[31] (the side chains of glutamate, aspatate, arginine, and lysine within 10 Å of the inhibitor, along with the amino and carboxy termini, were charged).[37] The total charge on the protein was neutral. All surface-exposed crystallographic water molecules were removed, while buried waters were maintained. The positions of all nonprotein heavy atoms, including all hydrogens and ligand atoms, were subjected to minimization with the DISCOVER program[38] using the CFF91 force field (all other atoms were constrained to their crystallographic positions). The van der Waals radii used in the SIP were derived using an internal dielectric of 1, thus ignoring the role of solute polarization (although it is treated implicitly in the parameterization of the SIP). Solute polarizability arises principally from two sources, the change in charge distribution upon transfer from the gas phase to solution, and large-amplitude motions in macromolecules. The latter are accounted for here by using an internal dielectric of 4 in the following calculations. FDPB calculations for the inhibitors used a border space of 20 Å with a resolution of 0.25 Å. For the protein, an initial coarse grid with a border space of 20 Å and resolution of 0.50 Å was used to provide boundary conditions (focusing) in subsequent calculations with a border space of 8 Å (a grid extent of 65 Å) and resolution of 0.25 Å.

Charges and radii for the inhibitors were calculated using the SIP procedure either by assembling parameters obtained for the amidinium cation and the various benzene derivatives, or for the intact inhibitors; these parameters are displayed in Figure 3. In assembling the inhibitors from two fragments, the charges on the redundant hydrogens were added to the carbon to which they were bonded. Protein parameters were assigned values given in Table IV. Contributions to the relative binding energies of the benzamidine inhibitors are presented in Table V. Energies are presented relative to *p*-aminobenzamidine, the strongest binding inhibitor. Only electrostatic contributions to these energies have been calculated (no nonelectrostatic components are included). Also presented in Table V are results from recent experimental[27] measurements of the relative free energies of binding.

For the *p*-H, *p*-CH_3, and *p*-F derivatives, the calculated relative binding free energy ($\Delta\Delta G_{12}$) was within 2 kJ mol^{-1} of the quoted limits of uncertainty of the experimental values when calculated with either set of charges and radii (columns A and B in Table V). Solvation energies calculated by either charge/radii set generally agreed to within a few kJ mol^{-1} with the exception of the *p*-OH derivative in which the difference was 21.4

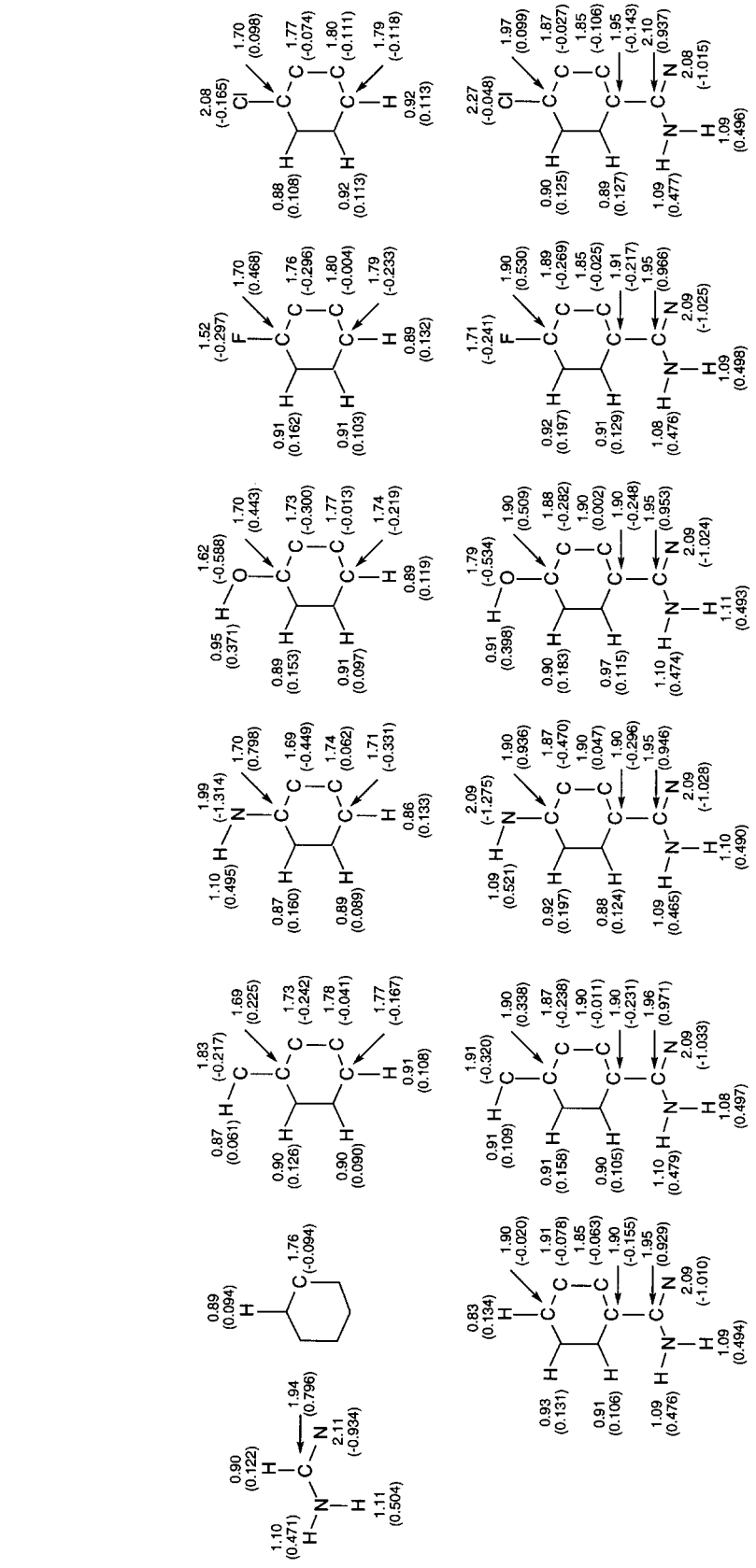

FIGURE 3. Atomic radii and charges (CHELPG, in parentheses) for the components of the benzamidine inhibitors calculated from the solvent interaction potential. Only the symmetry unique atoms are displayed. For toluene, the methyl hydrogen charges and radii are constrained to be equal; in all cases, the benzamidine charges and radii have C_{2v} symmetry constrained. CHELPG charges were calculated at the HF/6-31 + G(d)//MP2/6-31G(d) level.

TABLE V.
Comparison of Relative Binding Energies of Benzemidine Derivatives with Trypsin (kJ mol^{-1})[a].

	A[b]				B[c]				
	$\Delta\Delta G_{inter}$	$\Delta\Delta G_{solv}$(**L**)	$\Delta\Delta G_{solv}$(**L.P**)	$\Delta\Delta G_{12}$	$\Delta\Delta G_{inter}$	$\Delta\Delta G_{solv}$(**L**)	$\Delta\Delta G_{solv}$(**L.P**)	$\Delta\Delta G_{12}$	Expt.[d]
p-OH	−2.3	−12.9	−2.2	−13.0	−4.0	8.5	−2.7	7.2	1.3 ± 1.7
p-H	6.3	−1.9	3.1	1.3	3.7	2.7	3.4	2.9	1.7 ± 0.3
p-CH_3	6.9	−6.9	−1.0	1.0	4.5	−4.5	−1.5	1.5	2.8 ± 0.8
p-F	4.0	7.0	6.7	4.4	1.4	2.7	3.4	0.8	3.9 ± 1.3
p-Cl	4.3	9.4	1.5	12.1	2.3	3.1	−2.6	8.0	4.5 ± 2.5

[a] Energy differences are relative to *p*-aminobenzamidine. [b] Inhibitor charges and radii assembled from amidinium ion and benzene derivatives. [c] Inhibitor charges and radii from intact benzamidine deriatives. [d] Ref. 27.

kJ mol^{-1}. The *p*-OH derivative was incorrectly predicted to bind more tightly than the *p*-NH_2 derivative when the ''assembled'' charges and radii were used. The relative binding free energies for the *p*-Cl and *p*-OH derivatives were calculated to have the correct sign, although they were overestimated by 3.5 and 5.9 kJ mol^{-1}, respectively, when charges and radii for the intact inhibitor were used (column B). Thus, although *p*-aminobenzamidine was correctly predicted to bind most tightly, the binding energies of the other inhibitors relative to one another was incorrect.

It is well known that partial charges derived from the CHELPG method are not, in general, transferable to other molecules. This is highlighted by the large difference in inhibitor solvation energies calculated using the two charge/radii sets of roughly 25 kJ mol^{-1}. In comparison, the relative solvation energies seldom differ by more than 5 kJ mol^{-1}. Assembling charges and radii is most likely to lead to large errors in relative solvation energies in conjugated systems and where charge transfer between the fragments is likely.

The contribution of the solvation energy of the inhibitor–protein complex is nonnegligible and cannot be ignored. Omitting this contribution incorrectly predicts that the *p*-CH_3 and *p*-NH_2 derivatives bind with the same energy.

A recent FEP study[34] predicted relative binding free energies for the *p*-H, *p*-CH_3, and *p*-Cl derivatives of 8.4, 4.6, and 4.6 kJ mol^{-1}, respectively. In an earlier study,[32] a relative binding free energy of 3.8 kJ mol^{-1} was calculated for the *p*-F derivative, in excellent agreement with experiment.

Conclusions

The parameters that define the solvent interaction potential (SIP) are restricted to one van der Waals radius per atom and a single offset parameter. These are defined by the solvation energies (that are relatively well known) of a small set of systems that include both cations and anions. The solvation energies of the side chain molecules of the common amino acids are relatively well reproduced by the charges and radii derived using this approach. In particular, the solvation energies of the ionic systems are more accurately predicted using this method than with the PARSE parameters (the PARSE parameters were derived using experimental solvation energies that were significantly in error). For the neutral systems, the PARSE set performs significantly better. Using parameters derived using the SIP approach the relative binding energies of several benzamidine inhibitors for trypsin have been calculated. These calculations correctly predict that the *p*-NH_2 derivative is the stronger binding inhibitor when charge/radii parameters are derived for the intact inhibitor. For the *p*-H, *p*-CH_3, and *p*-F derivatives, the calculated relative binding free energies are within 2 kJ mol^{-1} of the quoted limits of uncertainty, irrespective of the charge/radii set used, whereas, for the *p*-Cl and *p*-OH derivatives, the error is 3.5 and 5.9 kJ mol^{-1}, respectively.

Acknowledgments

I wish to express my appreciation to Drs. V. Epa and N. E. Hall for helpful discussions relating to this work.

References

1. Sharp, K. A. Curr Opin Struct Biol 1994, 4, 234.
2. Orozco, M.; Alhambra, C.; Barril, X.; López, J. M.; Busquets, M. A.; Luque, F. J. J Mol Model 1996, 2, 1.

3. Tomasi, J.; Perisco, M. Chem Rev 1994, 94, 2027.
4. Sharp, K.; Honig, B. Annu Rev Biophys Biophys Chem 1990, 19, 301.
5. Rashin, A. A.; Namboodiri, K. J Phys Chem 1987, 91, 6003. Zauhar, R. J.; Morgan, R. S. J Comput Chem 1988, 9, 171. Juffer, A. H.; Botta, E. F. F.; van Keulen, B. A. M.; van der Ploeg, A.; Berendsen, H. J. C. J Comput Phys 1991, 97, 144. Rashin, A. A.; Young, L.; Topol, I. A. Biophys Chem 1994, 51, 359. Juffer, A. H.; Argosa, P.; Vogel, H. J. J Phys Chem 1997, 101, 7664.
6. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
7. Smith, B. J.; Hall, N. E. J Comput Chem 1998, 19, 1482.
8. The radii used to obtain γ and b in ref. 4 were not the radii determined here.
9. York, D. M.; Lee, T.-S.; Yang, W. Chem Phys Lett. 1996, 263, 297.
10. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab Initio Molecular Orbital Theory; Wiley: New York, 1986.
11. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, J. D.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. GAUSSIAN-94, Revision C.3, Gaussian, Inc., Pittsburgh, PA, 1995.
12. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. J Comput Chem 1993, 14, 1347.
13. Breneman, C. M.; Wiberg, K. B. J Comput Chem 1990, 3, 361.
14. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
15. DELPHI, Version 2.50; Biosym Technologies: San Diego, CA, 1993.
16. Nina, M.; Beglov, D.; Roux, B. J Phys Chem B 1997, 101, 5239.
17. (a) Bartmess, J. E., In NIST Standard Reference Database Number 69, Mallard, W. G.; Lindstrom, P. J., Eds; National Institute of Standards and Technology: Gaithersburg, MD, August 1997. (b) Hunter, E. P.; Lias, S. G. In NIST Standard Reference Database Number 69, Mallard, W. G.; Linstrom, P. J., Eds., National Institute of Standards and Technology: Gaithersburg, MD, August 1997 (http://webbook.nist.gov).
18. Hine, J.; Mookerjee, P. K. J Org Chem 1975, 40, 292.
19. Wagman, D. D.; Parker, V. B. P.; Schumm, R. H.; Halow, I.; Bailey, S. M.; Churney, K. L.; Nuttall, R. L. J Phys Chem Ref Data 1982, 11, Suppl. 2.
20. Lim, C.; Bashford, D.; Karplus, M. J Phys Chem 1991, 95, 5610.
21. Albert, A.; Sergeant, E. P. Ionization Constants of Acids and Bases; Metheun: London, 1962.
22. Pearson, R. G. J Am Chem Soc 1986, 108, 6109.
23. Maitland, G. C.; Rigby, M.; Smith, E. B.; Wakeham, W. A. Intermolecular Forces; Clarendon: Oxford, 1981.
24. Cramer, C. J.; Truhlar, D. G. J Comput-Aided Mol Des 1992, 6, 629.
25. Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. B. Biochemistry 1981, 20, 849.
26. Nina, M.; Beglov, D.; Roux, B. J Phys Chem B 1997, 101, 5239.
27. Mares-Guia, M.; Nelson, D. L.; Rogana, E. J Am Chem Soc 1977, 99, 2331.
28. Bode, W.; Schwager, P. J Mol Biol 1975, 98, 693.
29. Náray-Szabó, G. J Am Chem Soc 1984, 106, 4584.
30. Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. J Comput Chem 1995, 16, 454.
31. Åqvist, J. J Comput Chem 1996, 17, 1587.
32. Wong, C. F.; McCammon, J. A. J Am Chem Soc 1986, 108, 3830.
33. Resat, H.; Marrone, T. J.; McCammon, J. A. Biophys J 1997, 72, 522.
34. Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. J Phys Chem B 1997, 101, 9663.
35. Gilson, M. K.; Honig, B. Proteins 1988, 4, 7.
36. Takamatsu, Y.; Itai, A. Proteins 1998, 33, 62, and references therein.
37. Markley, J. L.; Porubcan, M. A. J Mol Biol 1976, 102, 487. Porubcan, M. A.; Neves, D. E.; Rausch, S. K.; Markley, J. L. Biochemistry 1978, 17, 4640. Sakiyama, F.; Kawata, Y. J Biochem 1983, 94, 1661.
38. DISCOVER 2.9.8, Molecular Simulations, San Diego, 1996. The charge on the amidinium carbon was set at +1.0.